Social Science and Humanities Journal
Everant Journal
ISSN: 2456-2653
Volume 09 Issue 01 January 2025
https://sshjournal.com/
Impact Factor: 2024: 6.576
2023: 5.731
DOI: https://doi.org/10.18535/sshj.v9i01.1575

Building Damage Detection Based on Earthquake Impact and Gradient Boosting Method

Elvan Felix Dwitama[1, a *] **| Genrawan Hoendarto**[2,b] **| Jimmy Tjen**[3,c]

[1,2,3]Widya Dharma Pontianak University, Pontianak, West Kalimantan, Indonesia

21421427_elvan_f_d@widyadharma.ac.id[1], genrawan@widyadharma.ac.id[2], jimmytjen@widyadharma.ac.id[3]

Received 05-12-2024
Revised 06-12-2024
Accepted 30-12-2024
Published 02-01-2025

Abstract:

Damage to buildings often occurs, and one of the causes is natural disaster such as earthquakes. Earthquakes frequently result in significant damage to buildings, causing financial losses due to damage to building facilities and even loss of life. Therefore, it is crucial to assess the damage to buildings to determine the extent of the damage. This research proposed an algorithm for detecting building damage using gradient boosting method. This method is similar to decision tree approach, but the decisions tree re-evaluated, resulting in smaller and more accurate data. For this analysis, the dataset was divided into two parts: training set and testing set. 80% of the dataset was used as training data, while 20% was used as testing data. After thorough data preprocessing, the gradient boosting method achieved an accuracy of 60.86% from large number of datasets compared to other methods, such as decision trees and random forests, the decision tree tends to overfit or underfit, especially with complex data. Meanwhile, the random forest method is generally faster and less prone to overfitting on large datasets. However, Gradient Boosting (GB) can achieve better accuracy, particularly for complex datasets. This result is indicating the effectiveness of the gradient boosting method. Despite the large and complex dataset, where prediction results can sometimes vary, the outcomes demonstrate good performance. Future research should focus on refining datasets and optimizing the parameters used for predicting building damage.

Keywords: Gradient Boosting, Earthquake Impact, Building Damage Detection.

Introduction:

Earthquake are among the most severe natural disasters, causing significant property damage and, more importantly, resulting in loss of life. Currently, accurately predicting the precise location where such events may occur remains a challenge [1]. Given that buildings constitute essential infrastructure utilized by many people, the accuracy and availability of information regarding structural damage are critical for effective post-disaster response and damage assessment [2]. This paper focuses on detecting building damage caused by earthquakes. The purpose of this journal is to detect the percentage level of damage that occurs in the building. This damage can include overall damage, damage based on the tilt of the building due to the earthquakes, damage that poses a risk to nearby buildings, severe damage to the building's foundation, moderate damage to the building's foundation, and minimal damage to the damage building's foundation [3]. This aims to determine the severity of the damage to the building through data from the building's structure [4]. The integration of hybrid deep learning models, as demonstrated by recent advancements, offers promising enhancements in the accuracy of detecting building damage from natural disasters like earthquakes [5]. Several studies have previously addressed the detection of building damage caused by earthquake using various approaches such as Convolutional Neural Networks (CNN) for building damage detection [6] and other using deep learning [7]

Previous research on building damage detection has been conducted by various authors. Author [8] detected building damage using an entropy-based sensor algorithm combined with machine learning and an auto regressive approach. The machine learning techniques used included Principal Component Analysis (PCA) and poly-exponential methods to develop a nonlinear model for detecting building damage, achieving prediction accuracy above 95%. The same author also employed a different approach [9], using an entropy-based sensor algorithm to detect building damage, with similar accuracy result exceeding 95%. Author [10], used OPCE algorithm to detect building damage, achieving accuracy of 83.3%, 97.4%, and 78.5% in cities of Yushu, Ishinomaki, and Mashiki, respectively. Meanwhile, author [11] used orthophoto imagery for building damage detection, with data collected from Kumamoto earthquake in Japan. The accuracy obtained in this study was 76.86%.

Another machine learning method which often used for regression task is the gradient boosting algorithm. The gradient boosting (GB) method is a machine learning method used to address regression and classification problems. This method can also utilize complex data to predict errors in the data. This method has been implemented in various studies. Author [12] used the gradient boosting method to predict short-term wind power output. From a fifteen-minute study, the test results achieved a performance with normalized mean absolute error of 5.15%. author [13] used the gradient boosting method to investigate anomalies in the electrical grid that cause data imbalance. According to author [13], the method was used because it demonstrated low performance values and had many variables. author [14] used gradient boosting to calculate energy usage in commercial building. The analysis results provided a prediction accuracy for R-squared and root mean square error (RMSE) of more than 80% from data tested on 410 commercial buildings. Gradient boosting has also been used to detect building damage. This research was conducted by author [15] and achieved high accuracy.

The Gradient boosting method operated by sequentially adding predictors to an ensemble, with each new predictor correcting the errors of its predecessor and fitting to the residuals left by previous predictors [16]. Gradient Boosting Decision Tree (GBDT) is widely used machine learning algorithm, with efficient implementations such as XG Boost and GBRT. Despite various engineering optimizations in these implementations, their efficiency and scalability are still limited when dealing with high-dimensional features and large datasets [17]. Gradient boosting is favored by many researchers due to its ability to deliver accurate results. For instance, researcher [18] utilized this method to predict household electricity consumption. Similarly, researcher [19] employed this method to forecast user preferences in online shopping.

In particular, this paper offers contribution as follows:

1. The implementation of GB algorithm in predicting earthquake impact occurred in the Gorkha district of Gandaki Pradesh, Nepal.
2. We propose new method detecting damage caused by earthquakes using GB method, which improves accuracy over traditional detection techniques.
3. This paper uses a large dataset of building structures affected by past earthquake, enabling more robust training and testing of the detection model.

4. The study compares the performance of GB with other machine learning techniques, such as Decision Trees and Random Forests, demonstrating its superior ability to handle complex datasets

This research is organized into four sections: the first section is the introduction, which explains the research object, the problems encountered, and the proposed solutions. The second section describes the gradient boosting method and discusses previous research findings using this machine learning technique.

Method:

This research proposed the use of gradient boosting to predict building damage. For a foundational understanding of the gradient boosting method and its previous predictive applications, please refer to [20] and [21] . The objective of our proposed method is to enhance the efficiency and accuracy of building damage predictions. The steps involved in the proposed method are outlined in Fig. 1. the primary method used for building damage detection is the GB algorithm. This technique was chosen due to its ability to handle complex, non-linear relationships between input features and building damage. GB iteratively improves the model by minimizing the residual error of previous models.

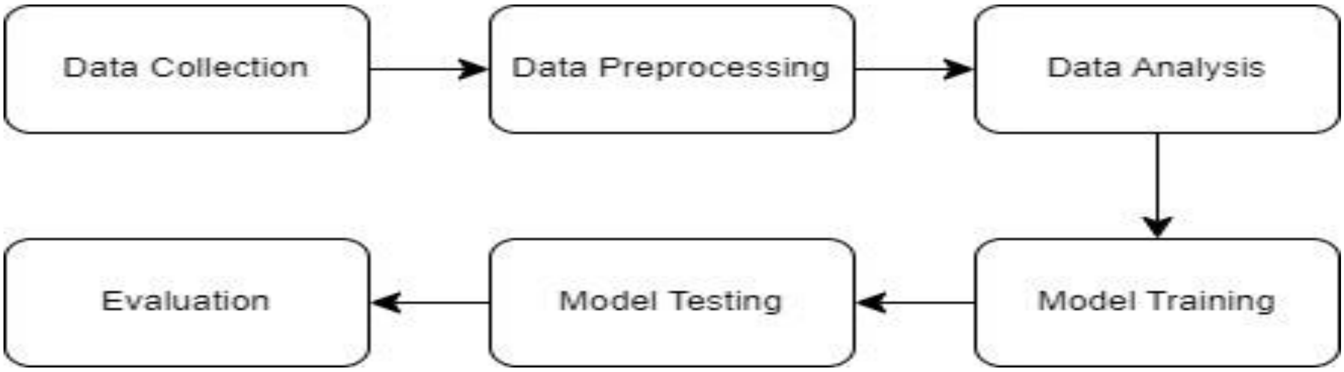

Fig. 1, Research Step

A. Data Collection

The dataset used in this research was obtained from Kaggle, which also provided the parameters for assessing building damage. The data utilized form Kaggle pertains to the earthquake that occurred in the Gorkha district of Gandaki Pradesh, Nepal. The dataset includes several parameters, such as building age, the number of floors before and after the earthquake, building height before and after quake, plinth area, and building condition after earthquake. This dataset offers a wide range variable for evaluating and predicting building damage. Such comprehensive data allows for a more accurate and detailed analysis of the factors contributing to structural damage.

B. Data Preprocessing

Initial preprocessing is a crucial step in preparing the dataset for machine learning models. The data preprocessing begins by prioritizing the handling of errors inherent in the method used. This approach aims to ensure that predictions achieve high accuracy and validity. This machine learning technique takes into account the building height before and after the earthquakes. This consideration is essential because some buildings experience no damage, while others suffer severe damage. The preprocessing steps is employed to ensure that the dataset is clean, and properly prepared for evaluation of machine learning models.

C. Data Analysis

The predictive model will be applied to the dataset, which includes the total number of samples. For this analysis, the dataset will be divided into two subnets: a training set and a testing set. The training set will use 80% of the total data to build the predictive model, while the remaining 20% will be used as the testing set to validate the model's accuracy. The analysis focused on determining the impact of various earthquake-related and building specific factors on damage severity. The importance of each features was evaluated using feature importance rankings provided by the GB model.

D. Model Training and Validation

Gradient boosting was chosen for its ability to handle complex data, including data that contains errors or inaccuracies. The model was trained using the training set on a system equipped with an 8-core processor, 16GB of RAM, and a graphics card with 2GB of VRAM.

E. Evaluation

After developing the building damage prediction model, it is crucial to assess the model's performance to determine its accuracy in predicting building damage. The model evaluation was conducted to assess the performance of gradient boosting. The model's performance was evaluated using a confusion matrix and accuracy metrics.

The confusion matrix is a tool used in classification problems to evaluate the performance of a classification model. Confusion matrix is a table that is used to define the performance of a classification algorithm Eq. 1. This matrix helps in assessing the accuracy and types of errors made by the model. The formula for this matrix is:

$$\text{Confusion matrix} = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} \quad (1)$$

Where the true represent the number of correct predictions and false the number of incorrect predictions. For the binary classification problem, the confusion matrix is typically structured as follows:

Table 1, Confusion matrix Structure

	Predictive Positive	Predictive Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Accuracy is derived from Confusion matrix and is expressed as Eq. 2.:

$$acc = \frac{1}{p}\sum_{i=1}^{p} n_y(i) \quad (2)$$

$$n_y(i) = \begin{cases} 1 \text{if } y(i) = \hat{y}(i) \\ 0 \text{ if otherwise} \end{cases}$$

Where n_y is a condition vector, which is 1 if the predicted and the real output is same, and vice versa. This metric is a type of average that's calculated times by 100% for the average accuracy percentage.

Result and Discussion:

In this section, we will present the result of the gradient boosting method. The result obtained from this method use classification. We will display a graph of the confusion matrix generated by calculating the datasets. This approach allows for a detailed evaluation of the model's performance across various classes within the dataset. By analyzing the confusion matrix, we can identify specific areas where the model excels or requires improvement. Previous studies have demonstrated the utility of confusion matrices in refining model parameters and enhancing overall predictive accuracy. The confusion matrix is a tool to evaluate the performance of a classification model by comparing the true labels with the predicted labels in this case, the matrix evaluates the ability of a Gradient Boosting Classifier to correctly classify buildings as **Damaged-Not used**, **Damaged-Repaired and used**, **and Not damage**. By providing a detailed overview of correct and incorrect prediction of each class, the matrix offers valuable insight into the model's strength and weakness. This analysis aids in identifying potential areas for improvement, such as class imbalances or tendencies to misclassify particular building damage categories

Class Definitions:

- **Damaged-Not used**: Buildings that were damaged and are no longer in use.

- **Damaged-Repaired and used**: Buildings that were damaged but have been repaired and are still in use
- **Not damaged**: Buildings that were not affected by the earthquake.

Key Observation:

1. True Positives (Diagonal elements):
   - **Damaged-Not used**: The model correctly classified 158,759 buildings in this category.
   - **Damaged-Repaired and used**: The model correctly predicted 1,428 buildings in this category.
   - **Not damaged**: The model correctly classified 23,318 buildings
2. False Positives (Off-diagonal elements):
   - **Damaged-Not used**: The model incorrectly classified 1,058 and 6,455 buildings as "Damaged-Not used" from the other categories.
   - **Damaged-Repaired and used**: The classifier mistakenly predicted 71,179 buildings that belong to this class and 13,768 from other categories.
   - **Not damaged**: The model incorrectly classified 24,389 and 1,175 buildings in this class from other categories

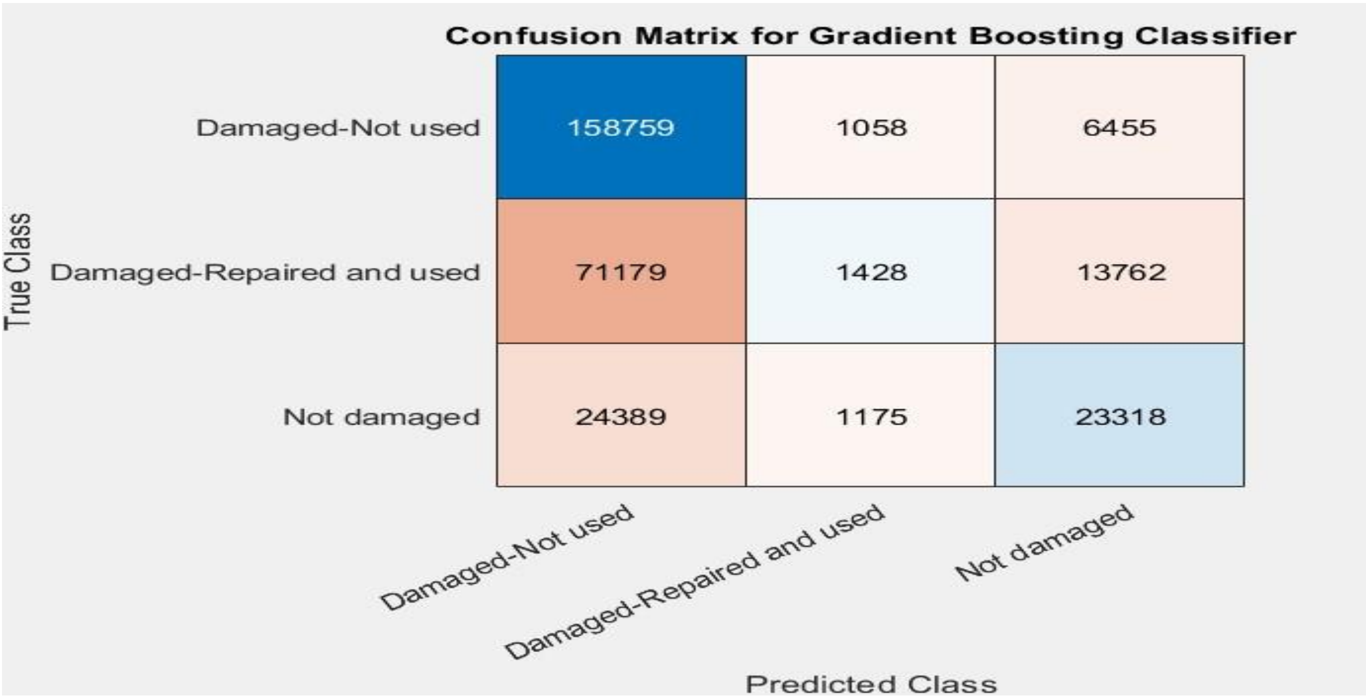

Fig. 2, Confusion Matrix for Gradient Boosting

Based on the Fig. 2, we calculated the accuracy obtained from the gradient boosting classification predictions. The result indicate that the model performs well under the given conditions, achieving a high level of precision in predicting building damage. These findings are consistent with previous studies that highlight the effectiveness of gradient boosting in handling complex datasets and improving prediction accuracy in various applications. The Gradient Boosting Classifier performs well for the **Damaged-Not used** class but exhibits some confusion between **Damaged-Repaired and used** and **Not damaged** categories. Further model tuning or feature engineering may improve the performance, especially in differentiating between repaired and still-used buildings and those that were not damaged. We also compared the accuracy of gradient boosting with the random forest with number of trees 100 and decision tree methods to differentiate the performance of these three approaches.

Table 2, Accuracy Result Using Confusion Matrix

Method	Accuracy
Gradient Boosting	60.86%
Random Forest (100)	59.37%
Decision Tree	55.64%

Table 2 provides evidence of the result obtained from the analysis using the three methods. The testing results show a high accuracy of 60.86%. Gradient boosting demonstrates a 1% higher accuracy compared to the random forest method and a 5% higher accuracy compared to the decision tree method. Among the three methods tested, we selected gradient boosting due to its ability to minimize prediction errors, particularly in complex datasets, resulting in higher accuracy.

Conclusion:

In this study, we employed the gradient boosting method to predict the extent of damage to buildings resulting from earthquakes. We utilized eight parameters to forecast the post-earthquake condition of the structures. The predictive model yielded an accuracy of 60.86% based on the results obtained from testing the gradient boosting algorithm. However, this study has certain limitations, particularly concerning in the inconsistency of the obtained accuracy. The accuracy tends to fluctuate due to insufficient data processing, although the variations are generally minor. Future research could focus on implementing more advance optimization techniques to enhance predictive accuracy. Additionally, applying this method to other datasets or incorporating a broader range of variables could potentially yield improved results. This analysis suggests the need for a focus on misclassified examples, particularly within the "Damaged-Repaired and used" class, for further model optimization.

References:

[1] M. Senthilkumar, D. Gnanasundar, B. Mohapatra, A. K. Jain, A. Nagar, and P. K. Parchure, "Earthquake prediction from high frequency groundwater level data: A case study from Gujarat, India," *HydroResearch*, vol. 3, pp. 118–123, Jan. 2020, doi: 10.1016/j.hydres.2020.10.004.

[2] M. Motosaka and K. Mitsuji, "Building damage during the 2011 off the Pacific coast of Tohoku Earthquake," *Soils Found.*, vol. 52, no. 5, pp. 929–944, 2012, doi: 10.1016/j.sandf.2012.11.012.

[3] A. Rao, J. Jung, V. Silva, G. Molinario, and S. H. Yun, "Earthquake building damage detection based on synthetic-aperture-radar imagery and machine learning," *Nat. Hazards Earth Syst. Sci.*, vol. 23, no. 2, pp. 789–807, Feb. 2023, doi: 10.5194/nhess-23-789-2023.

[4] H. Zheng *et al.*, "Assessment of Building Physical Vulnerability in Earthquake-Debris Flow Disaster Chain," *Int. J. Disaster Risk Sci.*, vol. 14, no. 4, pp. 666–679, Aug. 2023, doi: 10.1007/s13753-023-00509-7.

[5] P. K. E S, V. N. Thatha, G. Mamidisetti, S. V. Mantena, P. Chintamaneni, and R. Vatambeti, "Hybrid deep learning model with enhanced sunflower optimization for flood and earthquake detection," *Heliyon*, vol. 9, no. 10, 2023, doi: 10.1016/j.heliyon.2023.e21172.

[6] Y. Qing *et al.*, "Operational earthquake-induced building damage assessment using CNN-based direct remote sensing change detection on superpixel level," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 112, 2022, doi: 10.1016/j.jag.2022.102899.

[7] P. Mangalraj, S. P. Duddela, P. Kirubanantham, and S. Iniyan, "Post-earthquake Building Damage Detection Using Deep Learning," 2022, pp. 123–133. doi: 10.1007/978-981-16-1249-7_13.

[8] F. Smarra, J. Tjen, and A. D'Innocenzo, "Learning methods for structural damage

detection via entropy-based sensors selection," *Int. J. Robust Nonlinear Control*, vol. 32, no. 10, pp. 6035–6067, Jul. 2022, doi: 10.1002/rnc.6124.

[9] J. Tjen, G. Hoendarto, and T. Darmanto, "Ensemble of the Distance Correlation-Based and Entropy-Based Sensor Selection for Damage Detection," in *2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, Nov. 2022, pp. 183–189. doi: 10.1109/COMNETSAT56033.2022.9994387.

[10] Y. Nie, Q. Zeng, H. Zhang, and Q. Wang, "Building damage detection based on opce matching algorithm using a single post-event polsar data," *Remote Sens.*, vol. 13, no. 6, Mar. 2021, doi: 10.3390/rs13061146.

[11] B. Kalantar, N. Ueda, H. A. H. Al-Najjar, and A. A. Halin, "Assessment of convolutional neural network architectures for earthquake-induced building damage detection based on pre- and post-event orthophoto images," *Remote Sens.*, vol. 12, no. 21, 2020, doi: 10.3390/rs12213529.

[12] S. Park, S. Jung, J. Lee, and J. Hur, "A Short-Term Forecasting of Wind Power Outputs Based on Gradient Boosting Regression Tree Algorithms," *Energies*, vol. 16, no. 3, Feb. 2023, doi: 10.3390/en16031132.

[13] M. H. L. Louk and B. A. Tama, "Revisiting Gradient Boosting-Based Approaches for Learning Imbalanced Data: A Case of Anomaly Detection on Power Grids," *Big Data Cogn. Comput.*, vol. 6, no. 2, Jun. 2022, doi: 10.3390/bdcc6020041.

[14] S. Touzani, J. Granderson, and S. Fernandes, "Gradient boosting machine for modeling the energy consumption of commercial buildings," *Energy Build.*, vol. 158, pp. 1533–1543, Jan. 2018, doi: 10.1016/j.enbuild.2017.11.039.

[15] Y. He, H. Chen, D. Liu, and L. Zhang, "A framework of structural damage detection for civil structures using fast fourier transform and deep convolutional neural networks," *Appl. Sci.*, vol. 11, no. 19, Oct. 2021, doi: 10.3390/app11199345.

[16] G. Marasco *et al.*, "Machine learning approach to the safety assessment of a prestressed concrete railway bridge," *Struct. Infrastruct. Eng.*, vol. 20, no. 4, pp. 566–580, 2024, doi: 10.1080/15732479.2022.2119581.

[17] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, 2017, vol. 2017-December, pp. 3147–3155.

[18] P. Nie, M. Roccotelli, M. P. Fanti, Z. Ming, and Z. Li, "Prediction of home energy consumption based on gradient boosting regression tree," *Energy Reports*, vol. 7, pp. 1246–1255, Nov. 2021, doi: 10.1016/j.egyr.2021.02.006.

[19] Abdullah-All-Tanvir, I. Ali Khandokar, A. K. M. Muzahidul Islam, S. Islam, and S. Shatabda, "A gradient boosting classifier for purchase intention prediction of online shoppers," *Heliyon*, vol. 9, no. 4, Apr. 2023, doi: 10.1016/j.heliyon.2023.e15163.

[20] J. Ge, H. Tang, N. Yang, and Y. Hu, "Rapid identification of damaged buildings using incremental learning with transferred data from historical natural disaster cases," *ISPRS J. Photogramm. Remote Sens.*, vol. 195, pp. 105–128, Jan. 2023, doi: 10.1016/j.isprsjprs.2022.11.010.

[21] S. Ghimire and P. Guéguen, "Host-to-target region testing of machine learning models for seismic damage prediction in buildings," *Nat. Hazards*, vol. 120, no. 5, 2024, doi: 10.1007/s11069-023-06394-z.